DOCUMENT RESUME

ED 214 959 TM 820 110

TITLE	Evaluation Workshop V: Progress Evaluation. Team Notebook.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	73
NOTE	103p.; For related documents, see TM 820 106-111. Parts marginally legible.
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Administrators; Educational Assessment; Elementary Secondary Education; *Evaluation Methods; Evaluative Thinking; Evaluators; *Formative Evaluation; *Models; Training Methods; *Workshops

ABSTRACT

The handbook on progress evaluation was prepared for workshop participants. The Center for the Study of Evaluation (CSE) conducts a series of workshops for administrators and other evaluation team personnel. Specific exercises and activities are suggested using the five stage evaluation model. The phases are needs assessment, program planning, implementation, progress, and outcome evaluation. Participants are asked to categorize evaluation activities according to the five stage model. Implementation and progress evaluation activities are delineated. Evaluator trainees are assigned exercises in constructing affective measures. Evaluation measures should be selected according to an established set of criteria. The criteria must be ranked in order of relative importance. The logistics of organizing progress data are discussed. Results of evaluative measures must be gathered and presented in a report to decision makers. Practice in preparing these reports is part of the workshop training. This is done following a data analysis of results. This workshop is designed for detailed examination of the progress evaluation phase of the CSE evaluation model. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED214959

20

CSE
ARCHIVE COPY

Team No ________

COPY AVAILABLE

EVALUATION WORKSHOP V:
PROGRESS EVALUATION

TEAM NOTEBOOK

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G.S. Gray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM820 110

Copyright 1973, REGENTS OF THE UNIVERSITY OF CALIFORNIA

Copyright of these materials is claimed only during the period of development, test and evaluation unless authorization is granted by the National Institute of Education to claim copyright on the final materials. For information on the status of the copyright claim, contact either the copyright proprietor or the National Institute of Education.

CENTER FOR THE
STUDY OF
EVALUATION

MARVIN C. ALKIN
DIRECTOR

UCLA Graduate School of Education

The CENTER FOR THE STUDY OF EVALUATION is one of eight educational research and development centers sponsored by the National Institute of Education, Department of Health, Education and Welfare. Established at UCLA in 1966, under provisions of the Cooperative Research Act, CSE is devoted exclusively to the area of evaluation.

The mission of the Center is to conduct research and development activities for the production of new materials, practices and knowledges leading to the development of systems for evaluating education which can be adopted and implemented by educational agencies. The scope of activities includes the development of procedures and methodologies needed in the practical conduct of evaluation studies of various types, and the development of generalizable theories and concepts of evaluation relevant to different levels of education.

This publication is one of many produced by the Center toward its goals. Information on CSE and its publications may be obtained by writing to:

Dissemination Office
Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, California 90024

BEST COPY AVAILABLE

PART 1

Categorizing Evaluation Activities

Directions

On the next page you will find a partially completed diagram of the Center's five-stage model of evaluation. Your team's task is to categorize 15 specific evaluation activities in terms of their appropriateness to each stage.

The list of activities is contained in the envelope attached to this part of your notebook. Remove each activity from its self-adhesive backing and place it in the appropriate blank under each stage of the Center's model. You do not have to place the activities in sequence.

Two of the activities have already been categorized for you. Categorize the remaining 13 activities in the same manner.

BEST COPY AVAILABLE

Please turn the page.

Categorizing Evaluation Activities

NEEDS ASSESSMENT

Selecting a set of objectives from a list of instructional objectives

[Note: Remove adhesive backing from each evaluation activity and place it in the box representing the most appropriate evaluation stage]

PROGRAM PLANNING

IMPLEMENTATION EVALUATION	PROGRESS EVALUATION	OUTCOME EVALUATION
		Reviewing the results of the program in order to decide whether that program should be continued or dropped.

BEST COPY AVAILABLE

(continued)

PART 2

Please turn the page..

Categprizing Evaluation Activities - Feedback

Please insert this sheet at the end of Section C

NEEDS ASSESSMENT

Selecting a set of objectives from a list of instructional objectives

Establishing standards of performance to determine whether students are exceeding, achieving, or failing to reach proficiency on objectives for a new program that might be installed

Developing effective methods of combining the ratings of objectives which have been rated by a variety of individuals, e.g., parents, administrators, students, etc.

PROGRAM PLANNING

Developing procedures to investigate which of several kinds of instructional materials should be used in the program

Deciding upon the appropriate evaluation design to ascertain the effectiveness of the program

Deciding how the data about a program will be analyzed and reported

2-1
C-3

IMPLEMENTATION EVALUATION	PROGRESS EVALUATION	OUTCOME EVALUATION
Investigating whether students in a program are using the special equipment properly.	Preparing an interim report on whether the program is on schedule in meeting its instructional objectives	Reviewing the results of the program in order to decide whether that program should be continued or dropped
Observing a teacher's class to find out what he/she might be doing to account for the class' unusually high test scores	Establishing procedures for returning the students' scores on the weekly evaluations to the teachers within two days	Preparing a summary report on which students appeared to profit most from the new program
Using a checklist to determine whether the teachers followed the specially prepared lesson plans detailing the sequence for a given unit of instruction	Reporting the students' average score on each of the objectives of the program's first unit	Going back over the data collected during the course of the program to find out whether student interest in the subject matter increased, decreased, or remained the same

Please insert this sheet at the end of Section C

2-2
C-4

Please turn the page..

Categorizing Evaluation Activities - Feedback

Please insert this sheet at the end of Section C

NEEDS ASSESSMENT

Selecting a set of objectives from a list of instructional objectives

Establishing standards of performance to determine whether students are exceeding, achieving, or failing to reach proficiency on objectives for a new program that might be installed

Developing effective methods of combining the ratings of objectives which have been rated by a variety of individuals, e.g., parents, administrators, students, etc.

PROGRAM PLANNING

Developing procedures to investigate which of several kinds of instructional materials should be used in the program

Deciding upon the appropriate evaluation design to ascertain the effectiveness of the program

Deciding how the data about a program will be analyzed and reported

IMPLEMENTATION EVALUATION	PROGRESS EVALUATION	OUTCOME EVALUATION
Investigating whether students in a program are using the special equipment properly.	Preparing an interim report on whether the program is on schedule in meeting its instructional objectives	Reviewing the results of the program in order to decide whether that program should be continued or dropped
Observing a teacher's class to find out what he/she might be doing to account for the class' unusually high test scores	Establishing procedures for returning the students' scores on the weekly evaluations to the teachers within two days	Preparing a summary report on which students appeared to profit most from the new program
Using a checklist to determine whether the teachers followed the specially prepared lesson plans detailing the sequence for a given unit of instruction	Reporting the students' average score on each of the objectives of the program's first unit	Going back over the data collected during the course of the program to find out whether student interest in the subject matter increased, decreased, or remained the same

Please insert this sheet at the end of Section C

Please turn the page..

Categorizing Evaluation Activities - Feedback

Please insert this sheet at the end of Section C

NEEDS ASSESSMENT	PROGRAM PLANNING
Selecting a set of objectives from a list of instructional objectives	Developing procedures to investigate which of several kinds of instructional materials should be used in the program
Establishing standards of performance to determine whether students are exceeding, achieving, or failing to reach proficiency on objectives for a new program that might be installed	Deciding upon the appropriate evaluation design to ascertain the effectiveness of the program
Developing effective methods of combining the ratings of objectives which have been rated by a variety of individuals, e.g., parents, administrators, students, etc.	Deciding how the data about a program will be analyzed and reported

2-1
C-3

Please insert this sheet at the end of Section C

IMPLEMENTATION EVALUATION	PROGRESS EVALUATION	OUTCOME EVALUATION
Investigating whether students in a program are using the special equipment properly.	Preparing an interim report on whether the program is on schedule in meeting its instructional objectives	Reviewing the results of the program in order to decide whether that program should be continued or dropped
Observing a teacher's class to find out what he/she might be doing to account for the class' unusually high test scores	Establishing procedures for returning the students' scores on the weekly evaluations to the teachers within two days	Preparing a summary report on which students appeared to profit most from the new program
Using a checklist to determine whether the teachers followed the specially prepared lesson plans detailing the sequence for a given unit of instruction	Reporting the students' average score on each of the objectives of the program's first unit	Going back over the data collected during the course of the program to find out whether student interest in the subject matter increased, decreased, or remained the same

2-2
C-4

PART 3

Distinguishing Between Progress and Implementation Evaluation Activities

Directions

Circle your team's choice for each statement given on page F-1 using the following code:

I = Implementation Report
P = Progress Report
B = Both

You will have 10 minutes for this exercise.

1. I P B
2. I P B
3. I P B
4. I P B
5. I P B
6. I P B
7. I P B
8. I P B
9. I P B
10. I P B

STOP! PLEASE DO NOT TURN THE PAGE
UNTIL ASKED TO DO SO

CSE
ARCHIVE COPY

PART 4

Distinguishing Between Implementation and Progress Evaluation Activities - Feedback

1. I - The availability of textbooks is primarily relevant to the way the program is proceeding.

2. P - This statement refers to the extent to which the program's objectives are being achieved.

3. I - The biology teachers' contributions were to the operation of the program.

4. P - The rate at which students achieve objectives is a progress evaluation concern.

5. P - The objectives of the program may include those related to student attitudes.

6. B - Both implementation and progress evaluation are concerned with program modification.

7. I - The way in which students use instructional materials may deviate from the program plan. This is a procedural problem.

8. P - The number of students who achieve a program objective is one index of progress.

9. B - The superintendent needs implementation reports to explain progress results.

10. I (B) - The amount of time spent on homework is an index of the way in which the program is being run--except if doing homework is an objective of the program.

Please insert this sheet at the end of Section F.

4-1
F-2

Distinguishing Between Implementation and Progress Evaluation Activities - Feedback

1. I - The availability of textbooks is primarily relevant to the way the program is proceeding.

2. P - This statement refers to the extent to which the program's objectives are being achieved.

3. I - The biology teachers' contributions were to the operation of the program.

4. P - The rate at which students achieve objectives is a progress evaluation concern.

5. P - The objectives of the program may include those related to student attitudes.

6. B - Both implementation and progress evaluation are concerned with program modification.

7. I - The way in which students use instructional materials may deviate from the program plan. This is a procedural problem.

8. P - The number of students who achieve a program objective is one index of progress.

9. B - The superintendent needs implementation reports to explain progress results.

10. I (B) - The amount of time spent on homework is an index of the way in which the program is being run--except if doing homework is an objective of the program.

Please insert this sheet at the end of Section F.

Distinguishing Between Implementation and Progress Evaluation Activities - Feedback

1. I - The availability of textbooks is primarily relevant to the way the program is proceeding.

2. P - This statement refers to the extent to which the program's objectives are being achieved.

3. I - The biology teachers' contributions were to the operation of the program.

4. P - The rate at which students achieve objectives is a progress evaluation concern.

5. P - The objectives of the program may include those related to student attitudes.

6. B - Both implementation and progress evaluation are concerned with program modification.

7. I - The way in which students use instructional materials may deviate from the program plan. This is a procedural problem.

8. P - The number of students who achieve a program objective is one index of progress.

9. B - The superintendent needs implementation reports to explain progress results.

10. I (B) - The amount of time spent on homework is an index of the way in which the program is being run--except if doing homework is an objective of the program.

Please insert this sheet at the end of Section F.

PART 5

Selecting Progress Measures

Directions

On the following page is a partially completed memorandum to the Teacher Committee on Test Contruction from the progress evaluator, Ray Lopez. Your team's job is to complete Lopez' comments in the boxes provided.

You should refer to the test in Section G, and the articles in Sections I and J of your notebooks.

Please turn the page.

SAXONY UNIFIED SCHOOL DISTRICT

MEMORANDUM

November 3, 19XX

TO: The Teacher Committee on Test Construction

FROM: Ramon Lopez, Project Director

Subj: Selecting progress measures and administering tests

The following are my comments concerning the selection of the test items that were chosen to measure achievement of the objective, "to identify the causes and symptoms of venereal disease." I am also including a review of the procedures used to administer the test.

I. The Selection of Measures

Every item contains a flaw. I shall explain each by section and item number.

Section A

Line

1 Item One: This multiple choice item is supposedly testing
2 for the causes of gonorrhea. In fact, it is really testing
3 the student's ability to spell "gonococcus."
4 Item Two: Distractor D should be altered because
5
6
7 Item Three: This item does not provide students with
8 directions. Thus, some might respond to it by answering
9 "agree" or "disagree." Others could answer differently
10 because two other item formats which look like this one

Please turn the page

(continued)

are

Section B

Item One: If you look at the distractors, you will see that A or D has to be deleted because

Item Two: This item tends to instruct. The first sentence is unnecessary.

Item Three: The word "medicine" should be eliminated from the stem because it provides students with a

Also, the item is possibly an inappropriate measure for the objective because

Item Four: The use of the word "a" in the stem

Please turn the page

[redacted]

Item Five: All distractors in this item need revision because [redacted] Also, the item is not very good since [redacted]

Section C

Item One: This item has two problems: 1) 2) [redacted]

Item Two: [redacted]

Item Three: In writing a completion item, the blanks should be placed [redacted]

Please turn the page

(continued)

II. The Test Administration

39 The basic problem with the administration of the test
40 was that its procedures were not
41 Thus, we can be fairly certain that students did not
42 take the same tests. This casts doubts on the test's
43
44

45 The test was reproduced poorly. Too many items
46 were crowded on one page. Also, item 3, Section B,
47 and item 2, Section C, contained

48 I do not see the logic behind the different sections. Using sections
49 is usually only appropriate in a test where there are
50 several different item formats and separate directions
51 are needed for each. Also, the items were numbered
52 incorrectly because

53 certain rules which the test administrator must
54 adhere to, but which were violated are:
55 1. Avoid interruptions
56 2. Do not give hints or clues to
57 students concerning the correct answers.

STOP! PLEASE DO NOT TURN THE PAGE UNTIL ASKED TO DO SO

PART 6

SAXONY UNIFIED SCHOOL DISTRICT

MEMORANDUM

November 3, 19XX

TO: The Teacher Committee on Test Construction

FROM: Ramon Lopez, Project Director

Subj: Selecting progress measures and administering tests

The following are my comments concerning the selection of the test items that were chosen to measure achievement of the objective, "to identify the causes and symptoms of venereal disease." I am also including a review of the procedures used to administer the test.

I. The Selection of Measures

Every item contains a flaw. I shall explain each by section and item number.

Section A

Line

1 Item One: This multiple choice item is supposedly testing
2 for the causes of gonorrhea. In fact, it is really testing
3 the student's ability to spell "gonococcus."
4 Item Two: Distractor D should be altered because it serves
5 to increase the chances of getting the right answer for the
6 student who is willing to guess. (All distractors should be attractive to the uninformed student.)
7 Item Three: This item does not provide students with
8 directions. Thus, some might respond to it by answering
9 "agree" or "disagree." Others could answer differently
10 because two other item formats which look like this one

Please turn the page

(continued)

are

the true-false question and the essay.

Section B

Item One: If you look at the distractors, you will see that A or D has to be deleted because

both mean the same thing.

Item Two: This item tends to instruct. The first sentence is unnecessary.

Item Three: The word "medicine" should be eliminated from the stem because it provides students with a

hint or a clue.

Also, the item is possibly an inappropriate measure for the objective because

it is not consistent with it. The item deals with "treatment." However, it may be used to measure a related objective.

Item Four: The use of the word "a" in the stem

Please turn the page

gives away the answer since only distractor "D" fits.

Item Five: All distractors in this item need revision because they are of unequal size. Also, the item is not very good since it is not consistent with the objective. However, it may possibly be used to measure a related objective.

Section C

Item One: This item has two problems: 1) The distractors are of unequal size. 2) The item is not consistent with the objective. However, it might possibly be used to measure related objectives.

Item Two: This item is not necessarily consistent with the objective, although it might possibly be used to measure related objectives.

Item Three: In writing a completion item, the blanks should be placed at the end of the sentence.

Please turn the page

(continued)

II. The Test Administration

The basic problem with the administration of the test was that its procedures were not standardized.

Thus, we can be fairly certain that students did not take the same tests. This casts doubts on the test's reliability (consitency of measurement) and validity (accuracy of measurement).

The test was reproduced poorly. Too many items were crowded on one page. Also, item 3, Section B, and item 2, Section C, contained typographical errors.

I do not see the logic behind the different sections. Using sections is usually only appropriate in a test where there are several different item formats and separate directions are needed for each. Also, the items were numbered incorrectly because they were not consecutive.

certain rules which the test administrator must adhere to, but which were violated are:

1. Avoid interruptions
2. Do not give hints or clues to students concerning the correct answers.

Please insert this sheet at the end of Section H.

SAXONY UNIFIED SCHOOL DISTRICT

MEMORANDUM

November 3, 19XX

TO: The Teacher Committee on Test Construction

FROM: Ramon Lopez, Project Director

Subj: Selecting progress measures and administering tests

The following are my comments concerning the selection of the test items that were chosen to measure achievement of the objective, "to identify the causes and symptoms of venereal disease." I am also including a review of the procedures used to administer the test.

I. The Selection of Measures

Every item contains a flaw. I shall explain each by section and item number.

Section A

Line

1 Item One: This multiple choice item is supposedly testing
2 for the causes of gonorrhea. In fact, it is really testing
3 the student's ability to spell "gonococcus."
4 Item Two: Distractor D should be altered because it serves
5 to increase the chances of getting the right answer for the
6 student who is willing to guess. (All distractors should
be attractive to the uninformed student.)
7 Item Three: This item does not provide students with
8 directions. Thus, some might respond to it by answering
9 "agree" or "disagree." Others could answer differently
10 because two other item formats which look like this one

Please turn the page

(continued)

are the true-false question and the essay.

Section B

Item One: If you look at the distractors, you will see that A or D has to be deleted because both mean the same thing.

Item Two: This item tends to instruct. The first sentence is unnecessary.

Item Three: The word "medicine" should be eliminated from the stem because it provides students with a hint or a clue.

Also, the item is possibly an inappropriate measure for the objective because it is not consistent with it. The item deals with "treatment." However, it may be used to measure a related objective.

Item Four: The use of the word "a" in the stem

Please turn the page

gives away the answer since only distractor "D" fits.

Item Five: All distractors in this item need revision because they are of unequal size. Also, the item is not very good since it is not consistent with the objective. However, it may possibly be used to measure a related objective.

Section C

Item One: This item has two problems: 1) The distractors are of unequal size. 2) The item is not consistent with the objective. However, it might possibly be used to measure related objectives.

Item Two: This item is not necessarily consistent with the objective, although it might possibly be used to measure related objectives.

Item Three: In writing a completion item, the blanks should be placed at the end of the sentence.

Please turn the page

(continued)

II. The Test Administration

The basic problem with the administration of the test was that its procedures were not standardized. Thus, we can be fairly certain that students did not take the same tests. This casts doubts on the test's reliability (consitency of measurement) and validity (accuracy of measurement).

The test was reproduced poorly. Too many items were crowded on one page. Also, item 3, Section B, and item 2, Section C, contained typographical errors.

I do not see the logic behind the different sections. Using sections is usually only appropriate in a test where there are several different item formats and separate directions are needed for each. Also, the items were numbered incorrectly because they were not consecutive.

certain rules which the test administrator must adhere to, but which were violated are:

1. Avoid interruptions
2. Do not give hints or clues to students concerning the correct answers.

Please insert this sheet at the end of Section H.

SAXONY UNIFIED SCHOOL DISTRICT

MEMORANDUM

November 3, 19XX

TO: The Teacher Committee on Test Construction

FROM: Ramon Lopez, Project Director

Subj: Selecting progress measures and administering tests

The following are my comments concerning the selection of the test items that were chosen to measure achievement of the objective, "to identify the causes and symptoms of venereal disease." I am also including a review of the procedures used to administer the test.

I. The Selection of Measures

Every item contains a flaw. I shall explain each by section and item number.

Section A

Line

Item One: This multiple choice item is supposedly testing for the causes of gonorrhea. In fact, it is really testing the student's ability to spell "gonococcus."

Item Two: Distractor D should be altered because it serves to increase the chances of getting the right answer for the student who is willing to guess. (All distractors should be attractive to the uninformed student.)

Item Three: This item does not provide students with directions. Thus, some might respond to it by answering "agree" or "disagree." Others could answer differently because two other item formats which look like this one

Please turn the page

are the true-false question and the essay.

Section B

Item One: If you look at the distractors, you will see that A or D has to be deleted because both mean the same thing.

Item Two: This item tends to instruct. The first sentence is unnecessary.

Item Three: The word "medicine" should be eliminated from the stem because it provides students with a hint or a clue.

Also, the item is possibly an inappropriate measure for the objective because it is not consistent with it. The item deals with "treatment." However, it may be used to measure a related objective.

Item Four: The use of the word "a" in the stem

Please turn the page

(continued)

gives away the answer since only distractor "D" fits.

Item Five: All distractors in this item need revision because they are of unequal size. Also, the item is not very good since it is not consistent with the objective. However, it may possibly be used to measure a related objective.

Section C

Item One: This item has two problems: 1) The distractors are of unequal size. 2) The item is not consistent with the objective. However, it might possibly be used to measure related objectives.

Item Two: This item is not necessarily consistent with the objective, although it might possibly be used to measure related objectives.

Item Three: In writing a completion item, the blanks should be placed at the end of the sentence.

Please turn the page

(continued)

II. The Test Administration

The basic problem with the administration of the test was that its procedures were not standardized. Thus, we can be fairly certain that students did not take the same tests. This casts doubts on the test's reliability (consitency of measurement) and validity (accuracy of measurement).

The test was reproduced poorly. Too many items were crowded on one page. Also, item 3, Section B, and item 2, Section C, contained typographical errors.

I do not see the logic behind the different sections. Using sections is usually only appropriate in a test where there are several different item formats and separate directions are needed for each. Also, the items were numbered incorrectly because they were not consecutive.

certain rules which the test administrator must adhere to, but which were violated are:

1. Avoid interruptions
2. Do not give hints or clues to students concerning the correct answers.

Please insert this sheet at the end of Section H.

PART 7

Constructing Affective Measures

Directions

In the space below, construct a measure to assess performance on the following objective: "The student will appreciate how important it is to our society that people know about the causes, nature, treatment, and prevention of veneral disease."

Please keep in mind the four criteria for evaluating a measure described on page L-1 of your notebooks, since at the conclusion of this exercise your measure will be judged by another team.

The description of your measure should include the following:

a. its measurement format, e.g., ratings, multiple choice, performance task
b. a sample item or explanation of what the student is going to be doing
c. procedures for scoring, e.g., two raters using a 5-point scale
d. the type of score(s) to be derived from the measure relative to the objective, e.g., the average number of points assigned by both raters.

Description and Directions Measure: (Insert name of of format: ____________)

Scoring

STOP! PLEASE DO NOT TURN THE PAGE
UNTIL ASKED TO DO SO!

PART 8

Rating and Ranking Measures

Directions

Rank each of the four measures described in Section M of your individual notebooks and the other team's measure, (a total of five measures), on each of the four criteria listed below. If two or more measures are of equal quality on a given criterion, assign them tied ranks.

In order to maintain consistency across teams, please assign the higher numbered ranks to the better measures. That is, assign the rank of "5" to the best measure on a given criterion and a rank of "1" to the worst measure on this criterion.

Please record your team's ranks in the table below: Measures #1, #2, #3, and #4 are in Section M of your individual notebook; Measure #5 is the one constructed by another team.

	Measure Number				
Criterion	1	2	3	4	5
1. Precision of Quantitative Score					
2. Validity					
3. Reliability					
4. Cost and Administrative Efficiency					
Totals	-	-	-	-	-

Be sure to compute the total ranks for each measure

STOP!

PLEASE DO NOT TURN THE PAGE UNTIL ASKED TO DO SO!

PART 9

Rating and Ranking Measures - Feedback

The table below represents CSE's best estimate of the quality of each of the four measures presented in Section M of your individual notebooks. Since there are only four measures used in this feedback, the highest rank in the table is "4". The rationale for assigning ranks is presented beneath the table.

	Measure Number			
Criterion	1	2	3	4
1. Precision of Quantitative Score	3	3	1	3
2. Validity	4	1	2	3
3. Reliability	4	1	2	3
4. Cost and Administrative Efficiency	3	2	4	1
Totals	14	7	9	10

Rationale

1. Precision of Quantitative Score. All the measures except the performance task (Measure #3) yield a scale of scores having more than 2 points for each individual. No further distinctions are made in this exercise because of the subjective judgements involved in determining the precision of the various scales.

2. Validity. The interview results (Measure #2) are probably the least valid indicator of class performance since they may be biased both by the interviewer's perceptions and, more importantly, by the very few number of students interviewed per class. Measure #3, the performance task, is likely to be biased by friends tending to stick together on teams and/or students choosing to work with the brighter students on this task. The fourth measure, the content analysis of the essay, is almost as good as the questionnaire using ratings (Measure #1), but it is limited by the fact that it only examines one aspect of the objective, while the questionnaire apparently probes several facets of it.

3. Reliability. The reasons for assigning the ranks on this criterion are similiar to the ones for validity in that only three interviews per class would not provide a stable estimate of class performance on this objective, and the biases influencing the performance task might also lead to chance fluctuations (such as the brightest student deciding to take one position one day and a different position the next). The content analysis of the essay could be made more reliable by having several raters per paper plus several different essays per student. Since this is not the case in the present measure, it must be rated lower than the questionnaire.

4. Cost and Administrative Efficiency. The most expensive and time-consuming measure is the content analysis of the essay (Measure #4), since it involves a great deal of time and professional help in evaluating each paper. Greater efficiency would accrue from having two raters evaluate only a random sample of 25% of all the papers written (rather than having one rater for every paper). The interviews are also an expensive and time-consuming technique since they would probably require a minimum of 20 minutes per student and a professional interviewer would be needed to get good data. Unlike the essay, however, there is no simple way of improving the efficiency of this measure since there are too few students interviewed already and having more interviewers would only increase the costs. The questionnaire is, of course, a desireable technique from the point of view of efficiency since in only a few minutes all the students can be measured and no special administrator is needed. Further, the responses to short questionnaires using ratings like this one can be scored and tabulated quickly. The simplest and quickest measure is obviously the performance task.

Summary

Of the four measures discussed above, the questionnaire using ratings comes out on top. Although the essay and the performance task are tied for second, the essay is probably the better measure since its efficiency and reliability could be improved easily whereas this is not possible with the performance task. Even more important, however, is that the essay is likely to be more valid (since it is likely to be less biased by extraneous factors than is the performance task) and validity is generally a more important criterion than cost and efficiency. This latter point illustrates an important consideration in the evaluation of any measure, namely: Not only must one establish a set of criteria for selecting an evaluation measure, one must also decide the relative importance of each criterion in order to reach an appropriate decision for a given context.

Rating and Ranking Measures - Feedback

The table below represents CSE's best estimate of the quality of each of the four measures presented in Section M of your individual notebooks. Since there are only four measures used in this feedback, the highest rank in the table is "4". The rationale for assigning ranks is presented beneath the table.

	Measure Number			
Criterion	1	2	3	4
1. Precision of Quantitative Score	3	3	1	3
2. Validity	4	1	2	3
3. Reliability	4	1	2	3
4. Cost and Administrative Efficiency	3	2	4	1
Totals	14	7	9	10

Rationale

1. Precision of Quantitative Score. All the measures except the performance task (Measure #3) yield a scale of scores having more than 2 points for each individual. No further distinctions are made in this exercise because of the subjective judgements involved in determining the precision of the various scales.

2. Validity. The interview results (Measure #2) are probably the least valid indicator of class performance since they may be biased both by the interviewer's perceptions and, more importantly, by the very few number of students interviewed per class. Measure #3, the performance task, is likely to be biased by friends tending to stick together on teams and/or students choosing to work with the brighter students on this task. The fourth measure, the content analysis of the essay, is almost as good as the questionnaire using ratings (Measure #1), but it is limited by the fact that it only examines one aspect of the objective, while the questionnaire apparently probes several facets of it.

3. Reliability. The reasons for assigning the ranks on this criterion are similiar to the ones for validity in that only three interviews per class would not provide a stable estimate of class performance on this objective, and the biases influencing the performance task might also lead to chance fluctuations (such as the brightest student deciding to take one position one day and a different position the next). The content analysis of the essay could be made more reliable by having several raters per paper plus several different essays per student. Since this is not the case in the present measure, it must be rated lower than the questionnaire.

4. Cost and Administrative Efficiency. The most expensive and time-consuming measure is the content analysis of the essay (Measure #4), since it involves a great deal of time and professional help in evaluating each paper. Greater efficiency would accrue from having two raters evaluate only a random sample of 25% of all the papers written (rather than having one rater for every paper). The interviews are also an expensive and time-consuming technique since they would probably require a minimum of 20 minutes per student and a professional interviewer would be needed to get good data. Unlike the essay, however, there is no simple way of improving the efficiency of this measure since there are too few students interviewed already and having more interviewers would only increase the costs. The questionnaire is, of course, a desireable technique from the point of view of efficiency since in only a few minutes all the students can be measured and no special administrator is needed. Further, the responses to short questionnaires using ratings like this one can be scored and tabulated quickly. The simplest and quickest measure is obviously the performance task.

Summary

Of the four measures discussed above, the questionnaire using ratings comes out on top. Although the essay and the performance task are tied for second, the essay is probably the better measure since its efficiency and reliability could be improved easily whereas this is not possible with the performance task. Even more important, however, is that the essay is likely to be more valid (since it is likely to be less biased by extraneous factors than is the performance task) and validity is generally a more important criterion than cost and efficiency. This latter point illustrates an important consideration in the evaluation of any measure, namely: Not only must one establish a set of criteria for selecting an evaluation measure, one must also decide the relative importance of each criterion in order to reach an appropriate decision for a given context.

Rating and Ranking Measures - Feedback

The table below represents CSE's best estimate of the quality of each of the four measures presented in Section M of your individual notebooks. Since there are only four measures used in this feedback, the highest rank in the table is "4". The rationale for assigning ranks is presented beneath the table.

	Measure Number			
Criterion	1	2	3	4
1. Precision of Quantitative Score	3	3	1	3
2. Validity	4	1	2	3
3. Reliability	4	1	2	3
4. Cost and Administrative Efficiency	3	2	4	1
Totals	14	7	9	10

Rationale

1. Precision of Quantitative Score. All the measures except the performance task (Measure #3) yield a scale of scores having more than 2 points for each individual. No further distinctions are made in this exercise because of the subjective judgements involved in determining the precision of the various scales.

2. Validity. The interview results (Measure #2) are probably the least valid indicator of class performance since they may be biased both by the interviewer's perceptions and, more importantly, by the very few number of students interviewed per class. Measure #3, the performance task, is likely to be biased by friends tending to stick together on teams and/or students choosing to work with the brighter students on this task. The fourth measure, the content analysis of the essay, is almost as good as the questionnaire using ratings (Measure #1), but it is limited by the fact that it only examines one aspect of the objective, while the questionnaire apparently probes several facets of it.

3. Reliability. The reasons for assigning the ranks on this criterion are similiar to the ones for validity in that only three interviews per class would not provide a stable estimate of class performance on this objective, and the biases influencing the performance task might also lead to chance fluctuations (such as the brightest student deciding to take one position one day and a different position the next). The content analysis of the essay could be made more reliable by having several raters per paper plus several different essays per student. Since this is not the case in the present measure, it must be rated lower than the questionnaire.

4. Cost and Administrative Efficiency. The most expensive and time-consuming measure is the content analysis of the essay (Measure #4), since it involves a great deal of time and professional help in evaluating each paper. Greater efficiency would accrue from having two raters evaluate only a random sample of 25% of all the papers written (rather than having one rater for every paper). The interviews are also an expensive and time-consuming technique since they would probably require a minimum of 20 minutes per student and a professional interviewer would be needed to get good data. Unlike the essay, however, there is no simple way of improving the efficiency of this measure since there are too few students interviewed already and having more interviewers would only increase the costs. The questionnaire is, of course, a desireable technique from the point of view of efficiency since in only a few minutes all the students can be measured and no special administrator is needed. Further, the responses to short questionnaires using ratings like this one can be scored and tabulated quickly. The simplest and quickest measure is obviously the performance task.

Summary

Of the four measures discussed above, the questionnaire using ratings comes out on top. Although the essay and the performance task are tied for second, the essay is probably the better measure since its efficiency and reliability could be improved easily whereas this is not possible with the performance task. Even more important, however, is that the essay is likely to be more valid (since it is likely to be less biased by extraneous factors than is the performance task) and validity is generally a more important criterion than cost and efficiency. This latter point illustrates an important consideration in the evaluation of any measure, namely: Not only must one establish a set of criteria for selecting an evaluation measure, one must also decide the relative importance of each criterion in order to reach an appropriate decision for a given context.

PART 10

Memo on Logistics of Organizing Progress Data

Directions

Complete the following memorandum from Lopez to Murphy detailing the logistics involved in gathering and recording information from a quiz to measure a fourth objective of the Veneral Disease component of the Sex Education program. In filling in the blanks, you should select the principle or procedure that you believe would be the most effective, given the circumstances described in the memo.

Saxony Unified School District

Memorandum

To: Dr. Murphy
From: Ramon Lopez
Subj: Organizing Data from Analysis

Line no.

1 All 360 students in the eighth grade will take a ten-item, fifteen-
2 minute, true-false quiz to measure Objective 4 of the VD component
3 of the sex education program. Because the whole test and all the
4 answers can fit on one side of the page, the students will record
5 their answers on

Please turn the page

Scoring will be done by teachers and the project staff the day after testing and the results will be returned as soon as possible. Since students will be pencilling in, rather than circling, their answers, the scoring key will take the form of

This particular true-false quiz was developed in the Hamilton School District. A correlation of .81 was found between student scores from the first time it was administered to the same students a second time -- one week later. This testifies to the measure's

Various kinds of analyses will be performed on the quiz data. For example, the percentage of students who correctly answer each item will be computed. This type of analysis results in an index of

An item discrimination analysis will also be conducted. This involves

A number of special project record forms have been developed. For instance, the form for recording whether 75% of the students in the program have achieved Objective Four is illustrated in the figure

Please turn the page

(continued)

Figure

The data analysis will be done by computer. Thus, I have attached a card layout sheet. As you can see from it, in addition to punching information related to the student's background, e.g., ethnic affiliation, the total scores computed by the graders will be punched. Since a distractor analysis will also be conducted, keypunchers will punch

The cards will be both verified and

The following personnel will see to it that the tests are delivered on time, scored, recorded and analyzed: 1. Project management: will supervise appropriate data collection plan; will check data collection schedule. 2. Teachers: will administer tests. 3. Data collection supervisor: will check delivery of measures, test administration, and data collection. 4. The Data scoring supervisor will:

STOP! PLEASE DO NOT TURN THE PAGE
UNTIL TOLD TO DO SO.

PART 11

Memo on Logistics of Organizing Progress Data - Feedback

Saxony Unified School District

Memorandum

To: Dr. Murphy
From: Ramon Lopez
Subj: Organizing Data from Analysis

Line no.

1 All 360 students in the eighth grade will take a ten-item, fifteen-
2 minute, true-false quiz to measure Objective 4 of the VD component
3 of the sex education program. Because the whole test and all the
4 answers can fit on one side of the page, the students will record
5 their answers on their answer sheets directly,
6 Scoring will be done by teachers and the project staff the day
7 after testing and the results will be returned as soon as possible.
8 Since students will be pencilling in, rather than circling, their
9 answers, the scoring key will take the form of
10 a template (to make scoring more rapid).

11 This particular true-false quiz was developed in the Hamilton
12 School District. A correlation of .81 was found between student
13 scores from the first time it was administered to the same students

a second time -- one week later. This testifies to the measure's reliability.

Various kinds of analyses will be performed on the quiz data. For example, the percentage of students who correctly answer each item will be computed. This type of analysis results in an index of item difficulty.

An item discrimination analysis will also be conducted. This involves determining whether an item is effective in discriminating between those who are instructed in the program and those who are in the control group.

A number of special project record forms have been developed. For instance, the form for recording whether 75% of the students in the program have achieved Objective Four is illustrated in the figure

Figure
Achievement of Objective 4

Directions

Record the student's score and circle it if the achievement level is satisfactory.

Student ID# (or name) Score ____________

____

Total number of students with satisfactory achievement ____________

$$\frac{\text{No. achieving mastery}}{\text{No. of students in the program}} \times 100 = \text{percentage of students in the program with satisfactory achievement}$$

The data analysis will be done by computer. Thus, I have attached a card layout sheet. As you can see from it, in addition to punching information related to the student's background, e.g., ethnic affiliation, the total scores computed by the graders will be punched. Since a distractor analysis will also be conducted, keypunchers will

punch choices for each of the ten items.

The cards will be both verified and

interpreted.

The following personnel will see to it that the tests are delivered on time, scored, recorded and analyzed: 1. Project management: will supervise appropriate data collection plan; will check data collection schedule. 2. Teachers: will administer tests. 3. Data collection supervisor: will check delivery of measures, test administration, and data collection. 4. The Data Scoring supervisor will:

ensure that the plan for scoring and recording the data is carried out properly, including the distribution of papers to be scored and the rostering of information.

Memo on Logistics of Organizing Progress Data - Feedback

Saxony Unified School District

Memorandum

To: Dr. Murphy
From: Ramon Lopez
Subj: Organizing Data from Analysis

Line no.

1 All 360 students in the eighth grade will take a ten-item, fifteen-
2 minute, true-false quiz to measure Objective 4 of the VD component
3 of the sex education program. Because the whole test and all the
4 answers can fit on one side of the page, the students will record
5 their answers on their answer sheets directly.

6 Scoring will be done by teachers and the project staff the day
7 after testing and the results will be returned as soon as possible.
8 Since students will be pencilling in, rather than circling, their
9 answers, the scoring key will take the form of
10 a template (to make scoring more rapid).

11 This particular true-false quiz was developed in the Hamilton
12 School District. A correlation of .81 was found between student
13 scores from the first time it was administered to the same students

a second time -- one week later. This testifies to the measure's reliability.

Various kinds of analyses will be performed on the quiz data. For example, the percentage of students who correctly answer each item will be computed. This type of analysis results in an index of item difficulty.

An item discrimination analysis will also be conducted. This involves determining whether an item is effective in discriminating between those who are instructed in the program and those who are in the control group.

A number of special project record forms have been developed. For instance, the form for recording whether 75% of the students in the program have achieved Objective Four is illustrated in the figure

Figure
Achievement of Objective 4

Directions

Record the student's score and circle it if the achievement level is satisfactory.

Student ID# (or name) Score ____________

____

Total number of students with satisfactory achievement ________

$$\frac{\text{No. achieving mastery}}{\text{No. of students in the program}} \times 100 = \text{percentage of students in the program with satisfactory achievement}$$

The data analysis will be done by computer. Thus, I have attached a card layout sheet. As you can see from it, in addition to punching information related to the student's background, e.g., ethnic affiliation, the total scores computed by the graders will be punched. Since a distractor analysis will also be conducted, keypunchers will punch choices for each of the ten items. The cards will be both verified and interpreted.

The following personnel will see to it that the tests are delivered on time, scored, recorded and analyzed: 1. Project management: will supervise appropriate data collection plan; will check data collection schedule. 2. Teachers: will administer tests. 3. Data collection supervisor: will check delivery of measures, test administration, and data collection. 4. The Data scoring supervisor will: ensure that the plan for scoring and recording the data is carried out properly, including the distribution of papers to be scored and the rostering of information.

Memo on Logistics of Organizing Progress Data - Feedback

Saxony Unified School District

Memorandum

To: Dr. Murphy
From: Ramon Lopez
Subj: Organizing Data from Analysis

Line no.

1 All 360 students in the eighth grade will take a ten-item, fifteen-
2 minute, true-false quiz to measure Objective 4 of the VD component
3 of the sex education program. Because the whole test and all the
4 answers can fit on one side of the page, the students will record
5 their answers on their answer sheets directly,

6 Scoring will be done by teachers and the project staff the day
7 after testing and the results will be returned as soon as possible.
8 Since students will be pencilling in, rather than circling, their
9 answers, the scoring key will take the form of

10 a template (to make scoring more rapid).

11 This particular true-false quiz was developed in the Hamilton
12 School District. A correlation of .81 was found between student
13 scores from the first time it was administered to the same students

a second time -- one week later. This testifies to the measure's reliability.

Various kinds of analyses will be performed on the quiz data. For example, the percentage of students who correctly answer each item will be computed. This type of analysis results in an index of item difficulty.

An item discrimination analysis will also be conducted. This involves determining whether an item is effective in discriminating between those who are instructed in the program and those who are in the control group.

A number of special project record forms have been developed. For instance, the form for recording whether 75% of the students in the program have achieved Objective Four is illustrated in the figure

Figure
Achievement of Objective 4

Directions

Record the student's score and circle it if the achievement level is satisfactory.

Student ID# (or name)	Score

Total number of students with satisfactory achievement __________

$$\frac{\text{No. achieving mastery}}{\text{No. of students in the program}} \times 100 = \text{percentage of students in the program with satisfactory achievement}$$

The data analysis will be done by computer. Thus, I have attached a card layout sheet. As you can see from it, in addition to punching information related to the student's background, e.g., ethnic affiliation, the total scores computed by the graders will be punched. Since a distractor analysis will also be conducted, keypunchers will punch choices for each of the ten items.

The cards will be both verified and interpreted.

The following personnel will see to it that the tests are delivered on time, scored, recorded and analyzed: 1. Project management: will supervise appropriate data collection plan; will check data collection schedule. 2. Teachers: will administer tests. 3. Data collection supervisor: will check delivery of measures, test administration, and data collection. 4. The Data Scoring supervisor will:

ensure that the plan for scoring and recording the data is carried out properly, including the distribution of papers to be scored and the rostering of information.

CSE
ARCHIVE COPY

PART 12

Reporting to Decision Makers

Directions

Indicate which figure(s) below could illustrate each of the three methods described in Section Q. If none of the figures fits the five factors associated with a given method, construct a figure that would illustrate this method.

Figures

FIGURE #1

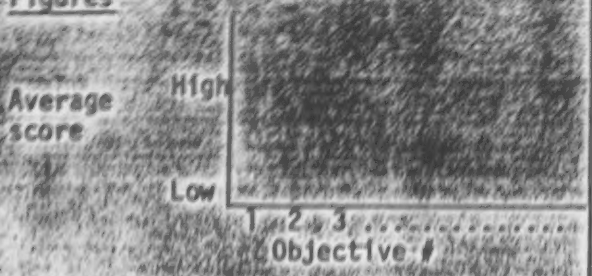


FIGURE #2

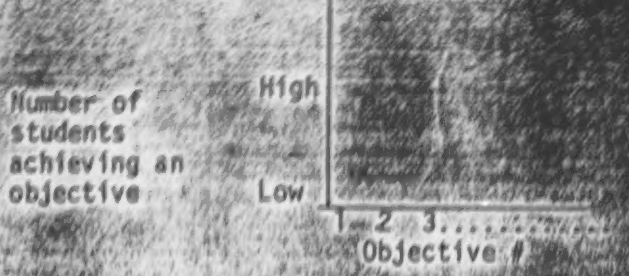


FIGURE #3

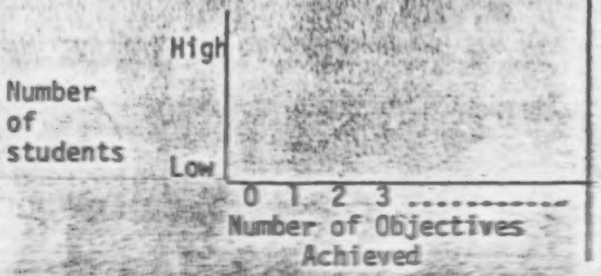


FIGURE #4

If necessary, use this space for your own figure. Be sure to label the two axes.

Questions

Circle the correct choice for each of the questions below:

A. Which figure(s) illustrate(s) Method A? 1 2 3 4

B. Which figure(s) illustrate(s) Method B? 1 2 3 4

C. Which figure(s) illustrate(s) Method C? 1 2 3 4

STOP! PLEASE DO NOT TURN THE PAGE
UNTIL ASKED TO DO SO

PART 13

Reporting to Decision Makers - Feedback

A with Figure #2
B with Figure #3
C with Figure #1

For your future reference, the three figures are illustrated below for the case of a program with ten objectives (five items per objective) and one hundred students:

FIGURE #1 (See Method C)

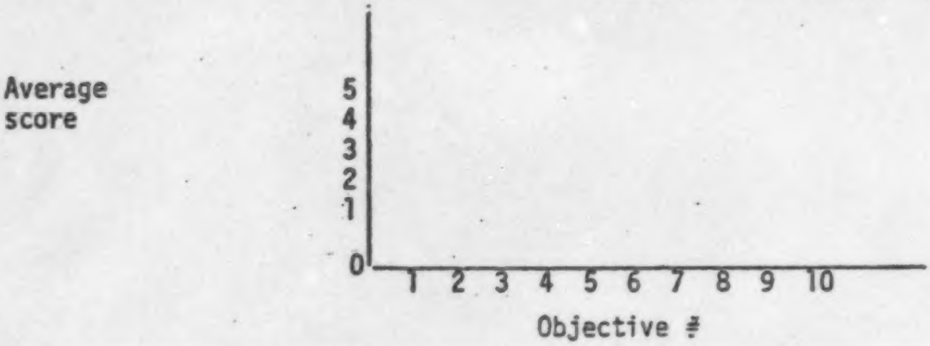


FIGURE #2 (See Method A)

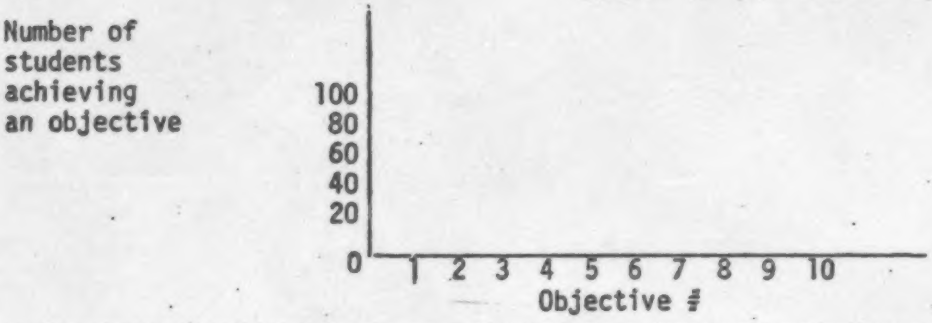
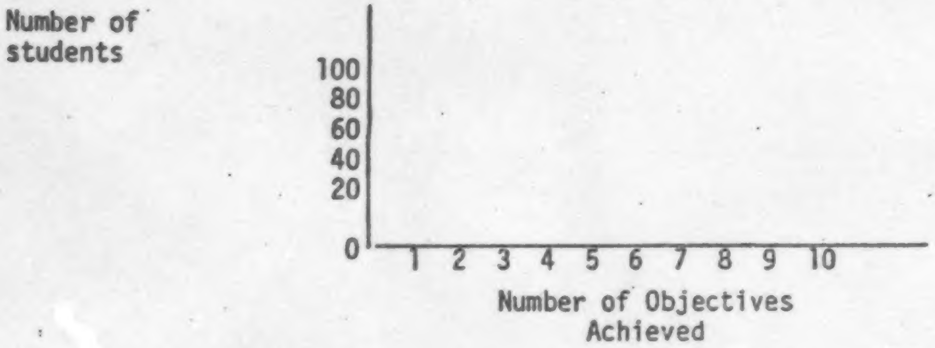


FIGURE #3 (See Method B)



Please insert this sheet at the end of Section Q.

Reporting to Decision Makers - Feedback

A with Figure #2
B with Figure #3
C with Figure #1

For your future reference, the three figures are illustrated below for the case of a program with ten objectives (five items per objective) and one hundred students:

FIGURE #1 (See Method C)

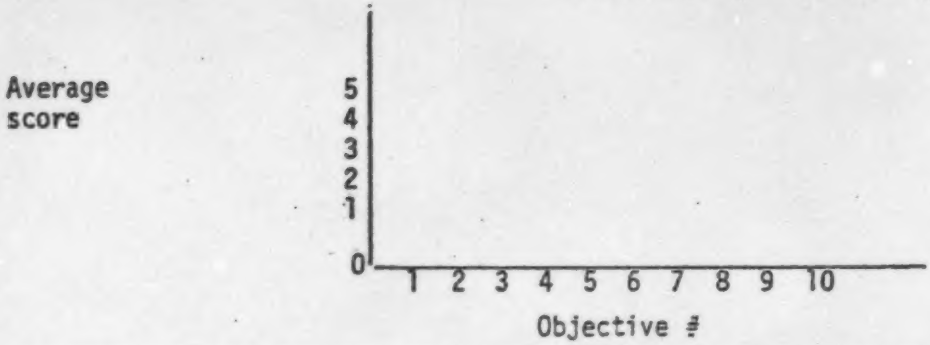


FIGURE #2 (See Method A)

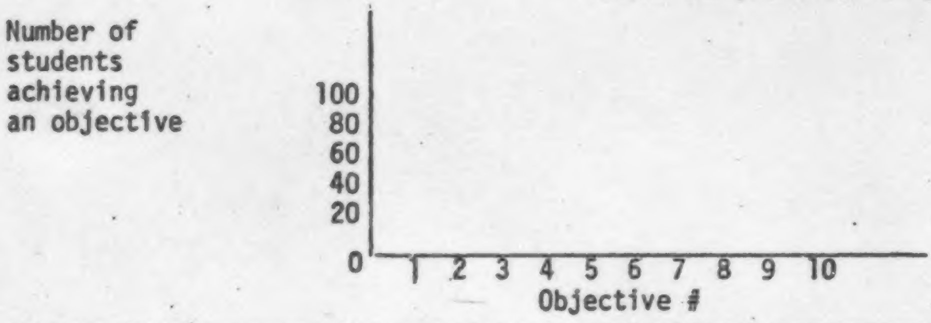
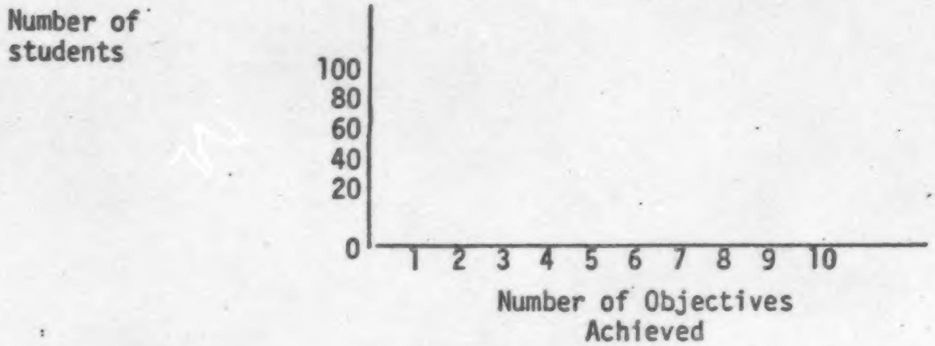


FIGURE #3 (See Method B)



Please insert this sheet at the end of Section Q.

Reporting to Decision Makers - Feedback

A with Figure #2
B with Figure #3
C with Figure #1

For your future reference, the three figures are illustrated below for the case of a program with ten objectives (five items per objective) and one hundred students:

FIGURE #1 (See Method C)

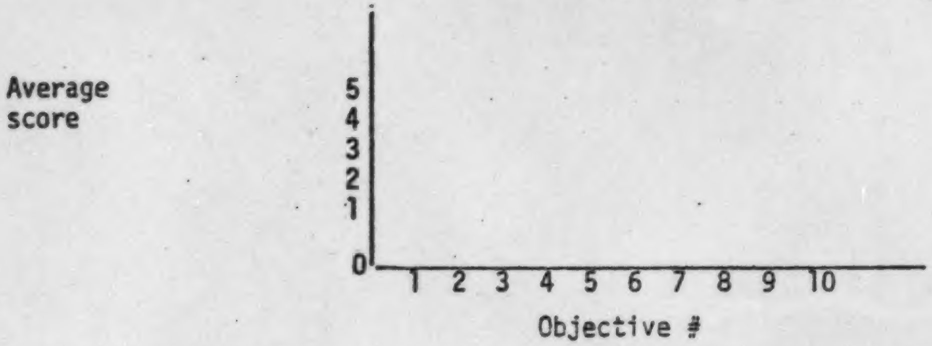


FIGURE #2 (See Method A)

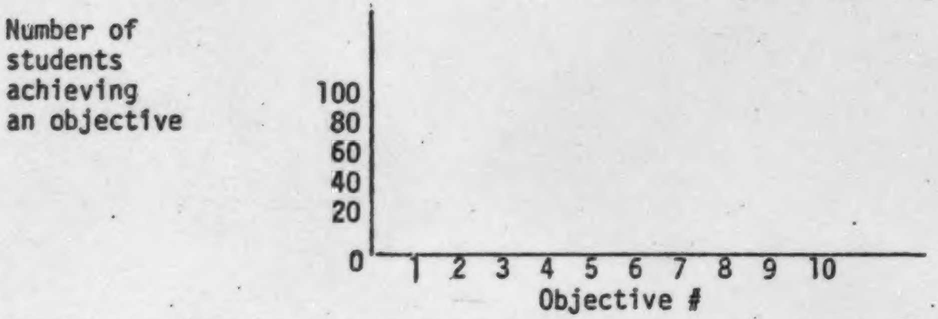
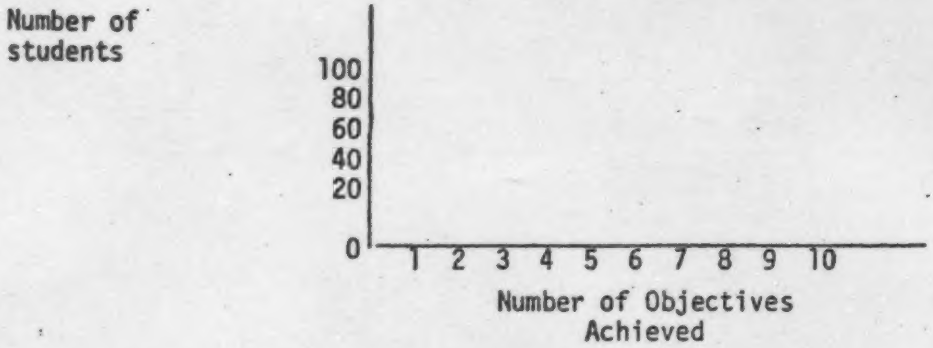


FIGURE #3 (See Method B)



Please insert this sheet at the end of Section Q.

PART 14

EVEN-NUMBERED TEAMS

Selecting Methods of Data Analysis

Directions

The table below contains the results obtained with a random sample of 10 ninth graders who took a test measuring the five objectives of the birth control unit. There were five items per objective. A perfect score was 25 items correct on the unit test.

Your task is to tabulate the results so that they look as good as possible. Extra columns and space have been provided for recording your answers. Please feel free to rearrange the data any way you wish.

Student	Objectives							
#	A	B	C	D	E			
1	5	4	5	3	4			
2	5	3	3	0	2			
3	5	5	4	2	4			
4	4	5	4	3	4			
5	5	4	4	3	4			
6	4	5	3	1	2			
7	5	5	5	5	5			
8	4	5	5	2	4			
9	5	4	2	5	4			
10	5	5	4	3	4			

Please use the other side of this page, if necessary.

PART 15

ODD-NUMBERED TEAMS

Selecting Methods of Data Analysis

Directions

The table below contains the results obtained with a random sample of 10 ninth graders who took a test measuring the five objectives of the birth control unit. There were five items per objective. A perfect score was 25 items correct on the unit test.

Your task is to tabulate the results so that they look as poor as possible. Extra columns and space have been provided for recording your answers. Please feel free to rearrange the data any way you wish.

Student	Objectives							
#	A	B	C	D	E			
1	5	4	5	3	4			
2	5	3	3	0	2			
3	5	5	4	2	4			
4	4	5	4	3	4			
5	5	4	4	3	4			
6	4	5	3	1	2			
7	5	5	5	5	5			
8	4	5	5	2	4			
9	5	4	2	5	4			
10	5	5	4	3	4			

Please use the other side of this page, if necessary.

PART 16

Selecting Methods of Data Analysis - Feedback

1. The progress results are good

Student #	Score on Objectives A	B	C	D	E	Total Score	Total Percent Correct on Test
1	5	4	5	3	4	21	84%
2	5	3	3	0	2	13	52%
3	5	5	4	2	4	20	80%
4	4	5	4	3	4	20	80%
5	5	4	4	3	4	20	80%
6	4	5	3	1	2	15	60%
7	5	5	5	5	5	25	100%
8	4	5	5	2	4	20	80%
9	5	4	2	5	4	20	80%
10	5	5	4	3	4	21	84%

Conclusion: Eight out of the ten students obtained a score of 80% or more.

2. The progress results are poor

Student #	Score on Objectives A	B	C	D	E	# of Objectives Mastered	Percentage of Objectives Mastered
1	5	4	5	3	4	2 out of 5	40%
2	5	3	3	0	2	1 out of 5	20%
3	5	5	4	2	4	2 out of 5	40%
4	4	5	4	3	4	1 out of 5	20%
5	5	4	4	3	4	1 out of 5	20%
6	4	5	3	1	2	1 out of 5	20%
7	5	5	5	5	5	5 out of 5	100%
8	4	5	5	2	4	2 out of 5	40%
9	5	4	2	5	4	2 out of 5	40%
10	5	5	4	3	4	2 out of 5	40%

Conclusion: Most students only mastered one or two objectives. Only one student mastered all objectives and thus, achieved more than 80%.

3. More specific progress results

Compute and then report the actual average score for each objective:

Objective	A	B	C	D	E
Average Score	4.7	4.5	4.1	2.7	3.7

Setting a mastery level is often an arbitrary decision that results in the omission of relevant information. For example, a level of "4" items correct for each objective might be more appropriate than "5," since it allows for chance errors. Using the actual average score for each objective (perhaps in conjunction with the number of pupils mastering it) is the best way of highlighting the existing trends in the data.

Please insert this sheet at the end of Section R or Section S.

Selecting Methods of Data Analysis - Feedback

1. The progress results are good

Student #	Score on Objectives A	B	C	D	E	Total Score	Total Percent Correct on Test
1	5	4	5	3	4	21	84%
2	5	3	3	0	2	13	52%
3	5	5	4	2	4	20	80%
4	4	5	4	3	4	20	80%
5	5	4	4	3	4	20	80%
6	4	5	3	1	2	15	60%
7	5	5	5	5	5	25	100%
8	4	5	5	2	4	20	80%
9	5	4	2	5	4	20	80%
10	5	5	4	3	4	21	84%

Conclusion: Eight out of the ten students obtained a score of 80% or more.

2. The progress results are poor

Student #	Score on Objectives A	B	C	D	E	# of Objectives Mastered	Percentage of Objectives Mastered
1	5	4	5	3	4	2 out of 5	40%
2	5	3	3	0	2	1 out of 5	20%
3	5	5	4	2	4	2 out of 5	40%
4	4	5	4	3	4	1 out of 5	20%
5	5	4	4	3	4	1 out of 5	20%
6	4	5	3	1	2	1 out of 5	20%
7	5	5	5	5	5	5 out of 5	100%
8	4	5	5	2	4	2 out of 5	40%
9	5	4	2	5	4	2 out of 5	40%
10	5	5	4	3	4	2 out of 5	40%

Conclusion: Most students only mastered one or two objectives. Only one student mastered all objectives and thus, achieved more than 80%.

3. More specific progress results

Compute and then report the actual average score for each objective:

Objective	A	B	C	D	E
Average Score	4.7	4.5	4.1	2.7	3.7

Setting a mastery level is often an arbitrary decision that results in the omission of relevant information. For example, a level of "4" items correct for each objective might be more appropriate than "5," since it allows for chance errors. Using the actual average score for each objective (perhaps in conjunction with the number of pupils mastering it) is the best way of highlighting the existing trends in the data.

Please insert this sheet at the end of Section R or Section S.

Selecting Methods of Data Analysis - Feedback

1. The progress results are good

Student #	Score on Objectives A	B	C	D	E	Total Score	Total Percent Correct on Test
1	5	4	5	3	4	21	84%
2	5	3	3	0	2	13	52%
3	5	5	4	2	4	20	80%
4	4	5	4	3	4	20	80%
5	5	4	4	3	4	20	80%
6	4	5	3	1	2	15	60%
7	5	5	5	5	5	25	100%
8	4	5	5	2	4	20	80%
9	5	4	2	5	4	20	80%
10	5	5	4	3	4	21	84%

Conclusion: Eight out of the ten students obtained a score of 80% or more.

2. The progress results are poor

Student #	Score on Objectives A	B	C	D	E	# of Objectives Mastered	Percentage of Objectives Mastered
1	5	4	5	3	4	2 out of 5	40%
2	5	3	3	0	2	1 out of 5	20%
3	5	5	4	2	4	2 out of 5	40%
4	4	5	4	3	4	1 out of 5	20%
5	5	4	4	3	4	1 out of 5	20%
6	4	5	3	1	2	1 out of 5	20%
7	5	5	5	5	5	5 out of 5	100%
8	4	5	5	2	4	2 out of 5	40%
9	5	4	2	5	4	2 out of 5	40%
10	5	5	4	3	4	2 out of 5	40%

Conclusion: Most students only mastered one or two objectives. Only one student mastered all objectives and thus, achieved more than 80%.

3. More specific progress results

Compute and then report the actual average score for each objective:

Objective	A	B	C	D	E
Average Score	4.7	4.5	4.1	2.7	3.7

Setting a mastery level is often an arbitrary decision that results in the omission of relevant information. For example, a level of "4" items correct for each objective might be more appropriate than "5," since it allows for chance errors. Using the actual average score for each objective (perhaps in conjunction with the number of pupils mastering it) is the best way of highlighting the existing trends in the data.

Please insert this sheet at the end of Section R or Section S.

PART 17

Analyzing Data and Reporting to Decision Makers

Directions

For each of the decisions listed below, select the data analysis and reporting method given on page U-2 that would provide the most useful information, and write your reasons for this choice (or reasons for using an alternative method with the present data.)

1. Should we permit further field testing in the Spring Semester?

____ Method A
____ Method B
____ Method C
____ Alternative Method

Reasons:

2. In case we field test in the Spring, how should I record student progress as a basis for expectations of the program?

____ Method A
____ Method B
____ Method C
____ Alternative Method

Reasons:

3. What components of the program need to be improved?

____ Method A
____ Method B
____ Method C
____ Alternative Method

Reasons:

4. Should I continue or dellay the field testing of the neew Birth Control Unit?

____Method A
____Method B
____Method C
____Alternative Method

Reasons:

____

____

____

STOP!! PLEASE DO NOT TURN THE PAGE UNTIL ASKED TO DO SO.

PART 18

Analyzing Data and Reporting to Decision Makers - Feedback

1. Should we permit further field testing in the Spring semester?

   Method B, since it provides a general index of overall performance, e.g., the number of students achieving at least 50% of the program's objectives.

2. In case we field test in the Spring, how should I record student progress as a basis for expectations of the program?

   Method A, since it provides precise information about student achievement levels in relation to each objective. Method C also provides useful information, but it is probably too general.

3. What components of the program need to be improved?

   Method A is best because it shows whether the program is leading toward achievement of each objective -- even though the goals may not yet have been reached and further curriculum modifications may be needed. Method A also has the advantage of indicating how close the program is moving toward the student's improvement or achievement of a preset standard. Method C only indicated whether or not the objectives have been achieved.

4. Should I continue or delay the field testing of the Birth Control Unit?

   Method C is best because it provides information about the number of students who achieve each objective. From this, the lab director could infer the usefulness of continuing with field testing. Method A would provide him/her with more precise information about performance on each objective and could be useful in helping him/her decide how close the students are to achieving them.

Please insert this sheet at the end of Section U

Analyzing Data and Reporting to Decision Makers - Feedback

1. Should we permit further field testing in the Spring semester?

   Method B, since it provides a general index of overall performance, e.g., the number of students achieving at least 50% of the program's objectives.

2. In case we field test in the Spring, how should I record student progress as a basis for expectations of the program?

   Method A, since it provides precise information about student achievement levels in relation to each objective. Method C also provides useful information, but it is probably too general.

3. What components of the program need to be improved?

   Method A is best because it shows whether the program is leading toward achievement of each objective -- even though the goals may not yet have been reached and further curriculum modifications may be needed. Method A also has the advantage of indicating how close the program is moving toward the student's improvement or achievement of a preset standard. Method C only indicated whether or not the objectives have been achieved.

4. Should I continue or delay the field testing of the Birth Control Unit?

   Method C is best because it provides information about the number of students who achieve each objective. From this, the lab director could infer the usefulness of continuing with field testing. Method A would provide him/her with more precise information about performance on each objective and could be useful in helping him/her decide how close the students are to achieving them.

Please insert this sheet at the end of Section U

Analyzing Data and Reporting to Decision Makers - Feedback

1. Should we permit further field testing in the Spring semester?

   Method B, since it provides a general index of overall performance, e.g., the number of students achieving at least 50% of the program's objectives.

2. In case we field test in the Spring, how should I record student progress as a basis for expectations of the program?

   Method A, since it provides precise information about student achievement levels in relation to each objective. Method C also provides useful information, but it is probably too general.

3. What components of the program need to be improved?

   Method A is best because it shows whether the program is leading toward achievement of each objective -- even though the goals may not yet have been reached and further curriculum modifications may be needed. Method A also has the advantage of indicating how close the program is moving toward the student's improvement or achievement of a preset standard. Method C only indicated whether or not the objectives have been achieved.

4. Should I continue or delay the field testing of the Birth Control Unit?

   Method C is best because it provides information about the number of students who achieve each objective. From this, the lab director could infer the usefulness of continuing with field testing. Method A would provide him/her with more precise information about performance on each objective and could be useful in helping him/her decide how close the students are to achieving them.

Please insert this sheet at the end of Section U

PART 19

Reporting to Decision Makers

Directions

Your team's task is to write the concluding section of the progress evaluation report given on pages W-2 to W-10. Essentially, this means answering the following two questions:

(1) To what extent is the unit achieving its objectives?

(2) Which component(s) of the unit need improvement?

Your answers to the questions should be based on the information provided in the progress evaluation report. You will have 20 minutes to complete the exercise.

Conclusions:

1. To what extent is the unit achieving its objectives?

2. Which component(s) of the unit need improvement?

Recommendations:

UNTIL ASKED TO DO SO!

PART 20

Reporting to Decision Makers - Feedback

Conclusions

1. To what extent is the unit achieving its objectives?

Students are generally achieving the first two objectives as evidenced by the significant increase in their average score on tests given before and after instruction, and by the number of students whose individual scores increased. The unit has also achieved the third objective as indicated by average and individual student performance on the attitude questionnaire, and the results of the interviews. However, increases in performance in relation to the fourth unit objective were not as large as increases for the other three objectives. The results of the questionnaire and interviews designed to measure achievement of the fourth objective were, however, positive.

2. What component(s) of the unit need improvement?

The component dealing with objective 4: the student will enjoy participation in the program. According to the questionnaire results students tended to be unsure about whether or not they would recommend a class similar to their own to their friends.

Recommendations

1. Because students performed well on three out of four objectives, and because their attitudes were generally favorable to the unit, the following recommendations, based primarily on the results of the interviews are being made:

   1) Check to see that teachers and students understand and have access to project personnel ot materials that describe and explain to them the purpose(s) of the lessons.

   2) Ensure that provisions are made so that all students have access to the instructional materials when needed. This may mean additional duplication, when possible, or sharing among classes in schools.

Please insert this sheet at the end of Section W.

Reporting to Decision Makers - Feedback

Conclusions

1. To what extent is the unit achieving its objectives?

Students are generally achieving the first two objectives as evidenced by the significant increase in their average score on tests given before and after instruction, and by the number of students whose individual scores increased. The unit has also achieved the third objective as indicated by average and individual student performance on the attitude questionnaire, and the results of the interviews. However, increases in performance in relation to the fourth unit objective were not as large as increases for the other three objectives. The results of the questionnaire and interviews designed to measure achievement of the fourth objective were, however, positive.

2. What component(s) of the unit need improvement?

The component dealing with objective 4: the student will enjoy participation in the program. According to the questionnaire results students tended to be unsure about whether or not they would recommend a class similar to their own to their friends.

Recommendations

1. Because students performed well on three out of four objectives, and because their attitudes were generally favorable to the unit, the following recommendations, based primarily on the results of the interviews are being made:

   1) Check to see that teachers and students understand and have access to project personnel ot materials that describe and explain to them the purpose(s) of the lessons.

   2) Ensure that provisions are made so that all students have access to the instructional materials when needed. This may mean additional duplication, when possible, or sharing among classes in schools.

Please insert this sheet at the end of Section W.

Reporting to Decision Makers - Feedback

Conclusions

1. To what extent is the unit achieving its objectives?

Students are generally achieving the first two objectives as evidenced by the significant increase in their average score on tests given before and after instruction, and by the number of students whose individual scores increased. The unit has also achieved the third objective as indicated by average and individual student performance on the attitude questionnaire, and the results of the interviews. However, increases in performance in relation to the fourth unit objective were not as large as increases for the other three objectives. The results of the questionnaire and interviews designed to measure achievement of the fourth objective were, however, positive.

2. What component(s) of the unit need improvement?

The component dealing with objective 4: the student will enjoy participation in the program. According to the questionnaire results students tended to be unsure about whether or not they would recommend a class similar to their own to their friends.

Recommendations

1. Because students performed well on three out of four objectives, and because their attitudes were generally favorable to the unit, the following recommendations, based primarily on the results of the interviews are being made:

   1) Check to see that teachers and students understand and have access to project personnel ot materials that describe and explain to them the purpose(s) of the lessons.

   2) Ensure that provisions are made so that all students have access to the instructional materials when needed. This may mean additional duplication, when possible, or sharing among classes in schools.

Please insert this sheet at the end of Section W.